Coursera Applied Data Science Capstone Project: Effectiveness of Physical Distancing in Slowing the Spread of the Covid 19 Coronavirus in Affected Countries.

Tom Cole

April 2020

# Introduction

At present the world is afflicted by a pandemic caused by the Covid 19 coronavirus. Different countries have taken different measures in response. Most of those measures are a combination of testing and physical distancing in order to identify Covid 19 infections so that the infected can be isolated and treated before they can spread the infection further.

The original outbreak occurred in China and they initiated a strict regime of physical distancing that had a high degree of compliance. Approximately three months after the first cases were recorded China began recording a slowing of the rate of new infections. Was that due isolating the community or could it simply be the natural arc of the epidemic?

Italy implemented physical distancing but with very low initial compliance and saw Covid 19 infections rise unabated. South Korea initiated an intense regime of testing and did not close down businesses and schools. Singapore took a similar approach.

Other countries have delayed physical distancing measures with seemingly no change to the infection curve.

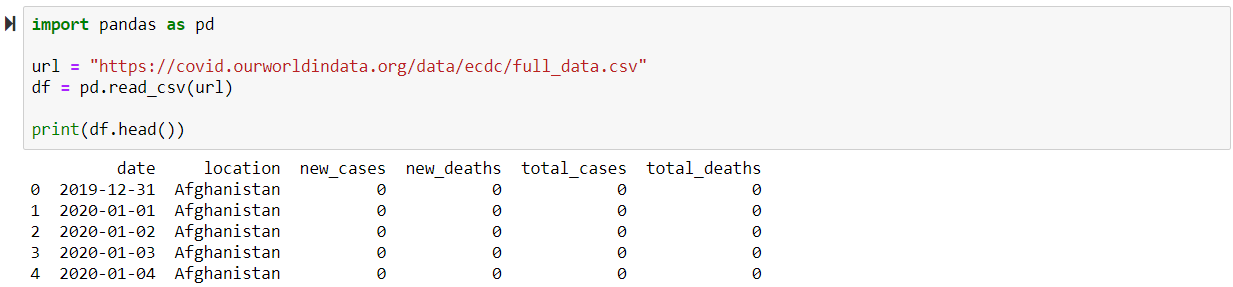
With governments and local communities debating the economic costs of physical distancing and "lockdowns" against the public health benefits can we determine the effectiveness of these measures from the currently available data?

# Data

Recovery and mortality rates will not be included as are not pertinent to the question. While they are directly related to infection, they are a function of the health system in the country and the underlying health of the infected individual and the population.

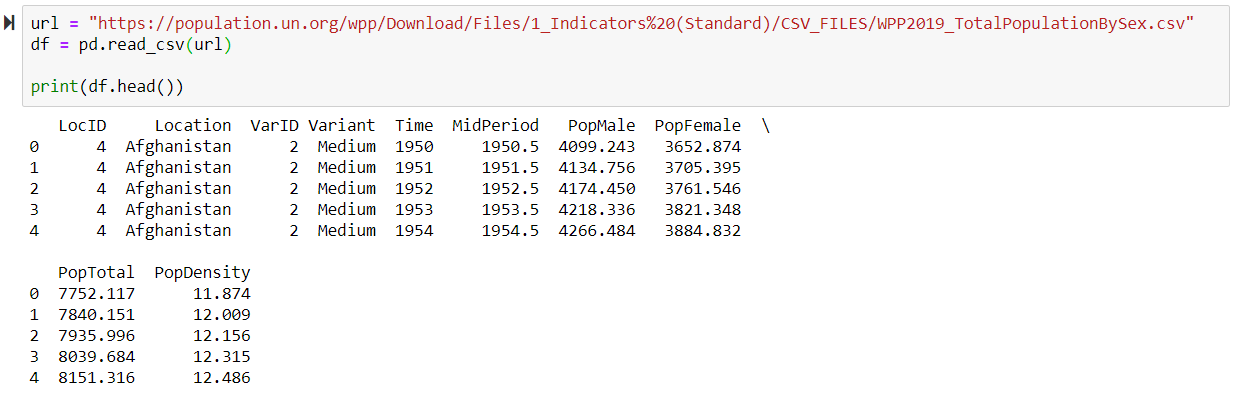
## Covid 19 Infection Data

Covid 19 Infection Data is sourced from the European Centre for Disease Prevention and Control and published on the website <https://ourworldindata.org/coronavirus-source-data> . This data includes the date of first case detection, the daily rate of new cases, the recovery rate, and the mortality rate for each country.



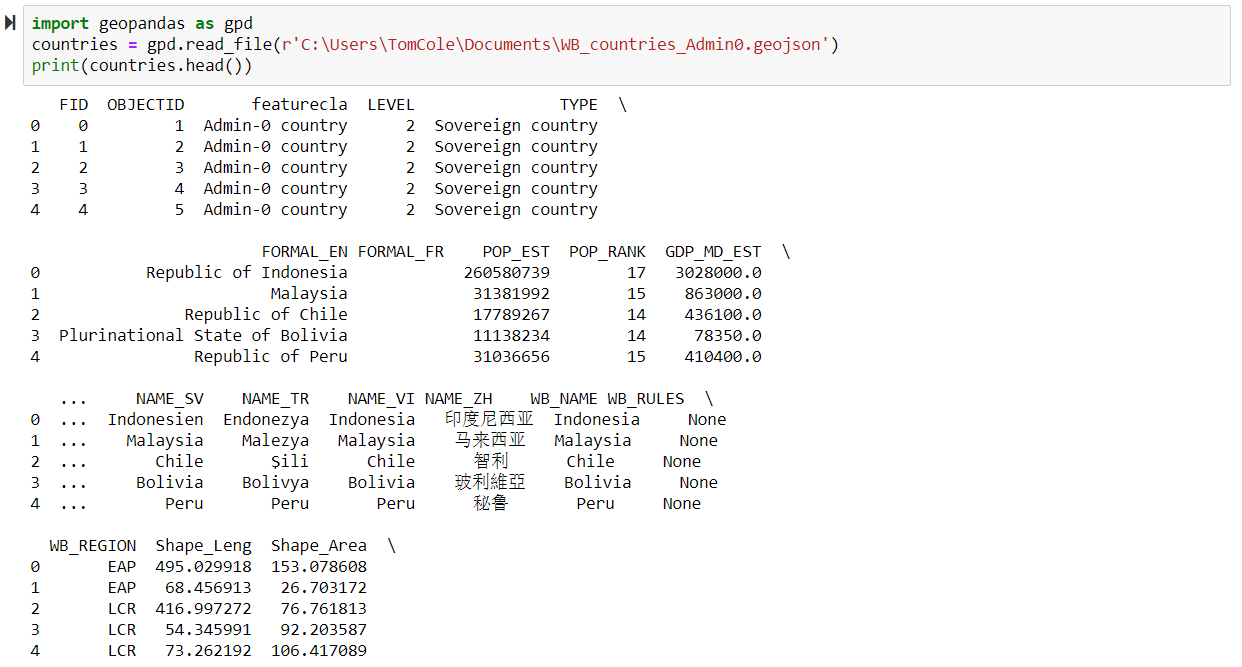
## 2.2 Country Population Data

Country population data to allow analysis of the infection rate as a percentage of population is sourced from https://population.un.org .



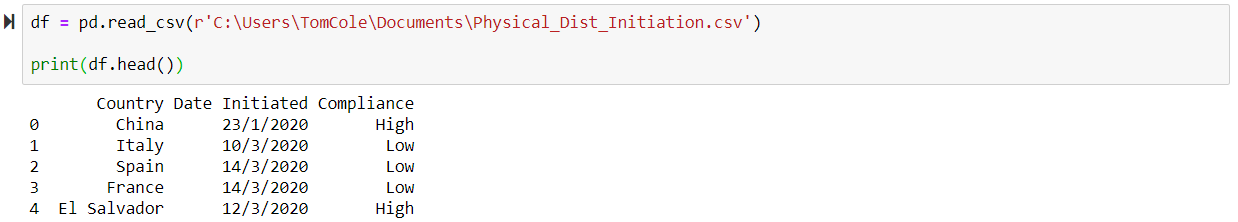
## 2.3 Country Geospatial Data

Country geospatial data to allow visualisation of infection rates is sourced from <https://datacatalog.worldbank.org/dataset/world-bank-official-boundaries>



## 2.4 Date of Initiation of Physical Distancing Measures

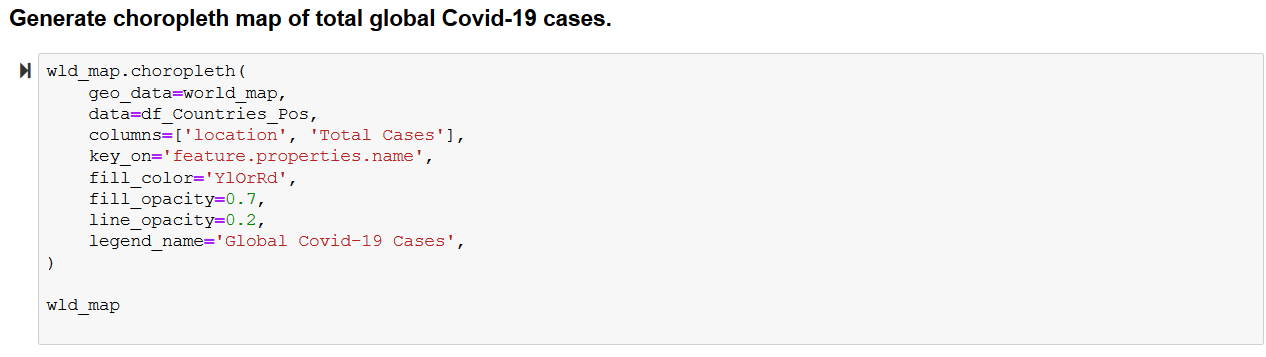
From media reports and government information we can establish when specific measures were enacted, particularly if and when they implemented physical distancing rules and what was the initial degree of compliance. Using infection rates after that date we should be able to identify if the measure is correlated to a change in infection rates.

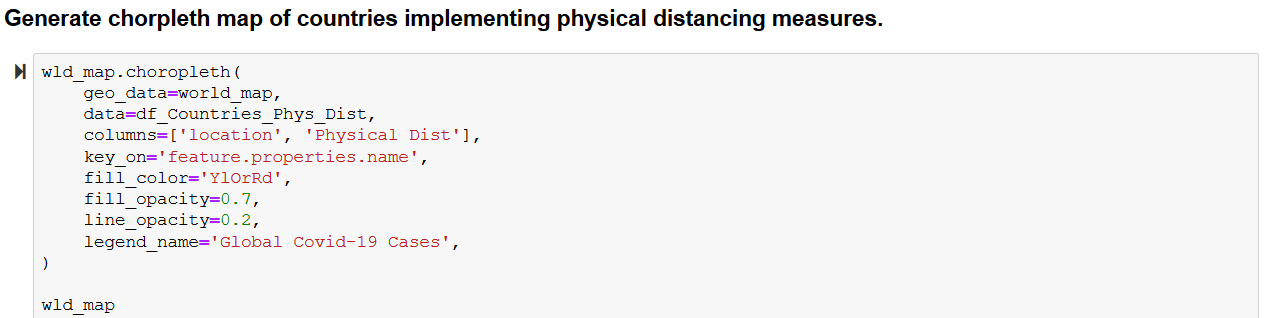


# Methodology

Generate maps to show global location of countries with corona virus and who have implemented physical distancing measures.

## Mapping With Folium





## Run Regression On key Features of The Data

Population Density on Total Cases



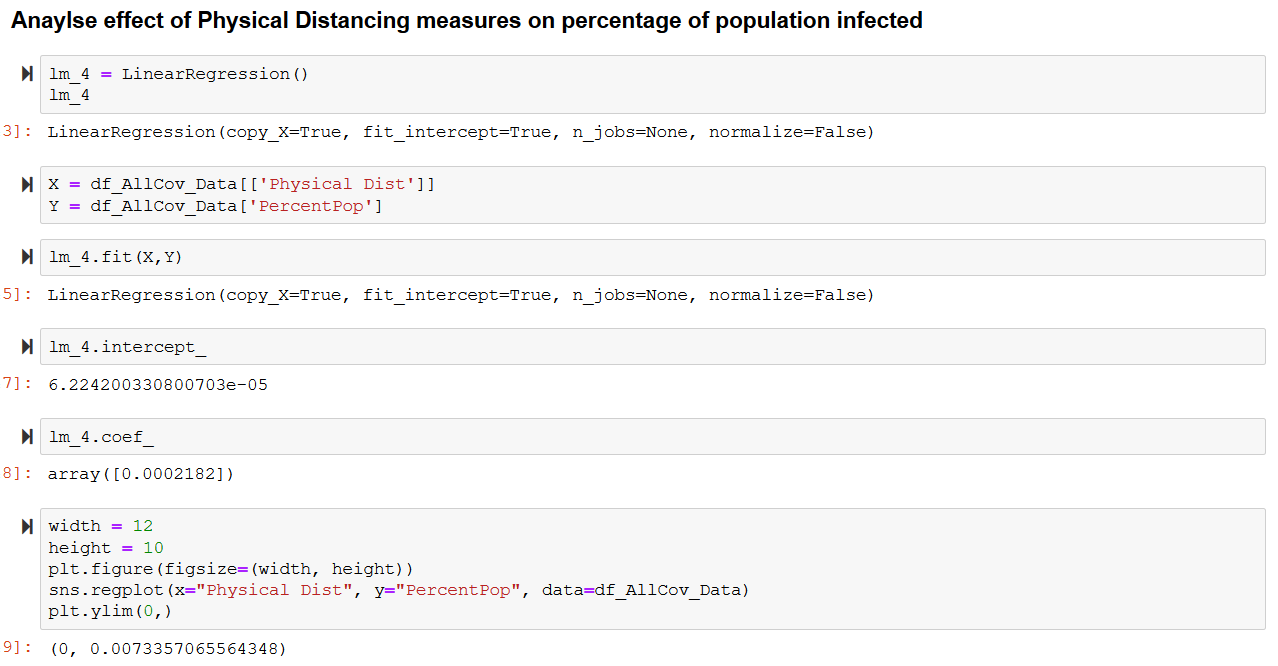
Population density on percentage of population infected.



Physical distancing on total cases



Physical Distancing on percentage population infected.



# Results

Analysis shows that there is a correlation between physical distancing measure being implemented and the total number of cases recorded in a country. This shows countries that have high numbers of cases are more likely to implement physical distancing measures.

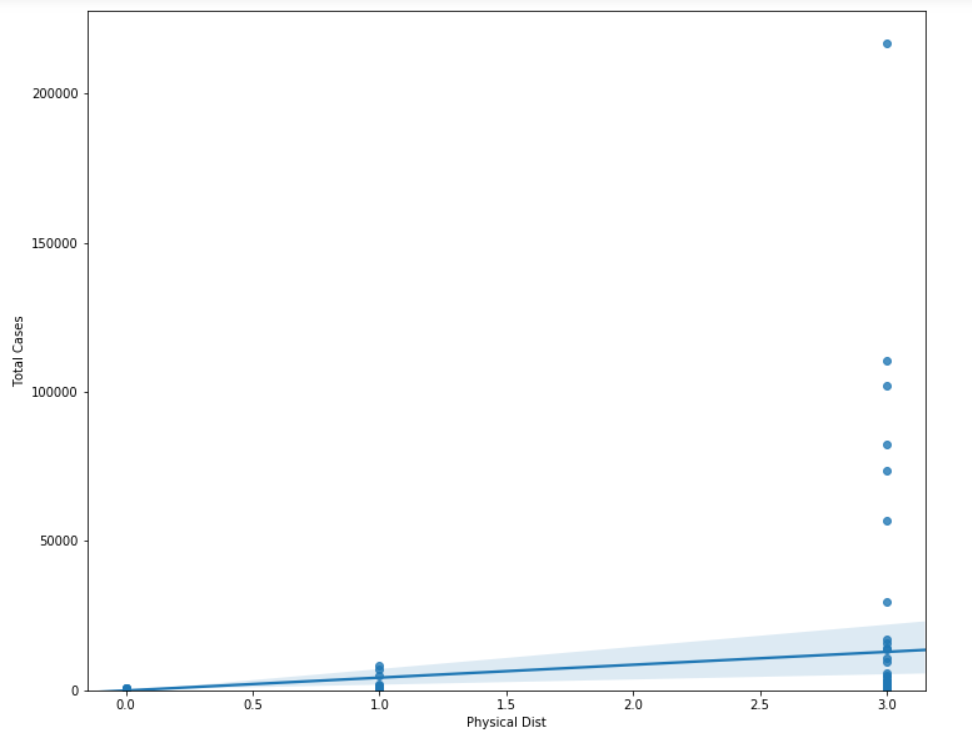


Figure 0=Unknown, 1= No Phys Distancing, 3=Phys Dist Implemented

A similar pattern is seen for the percentage of population

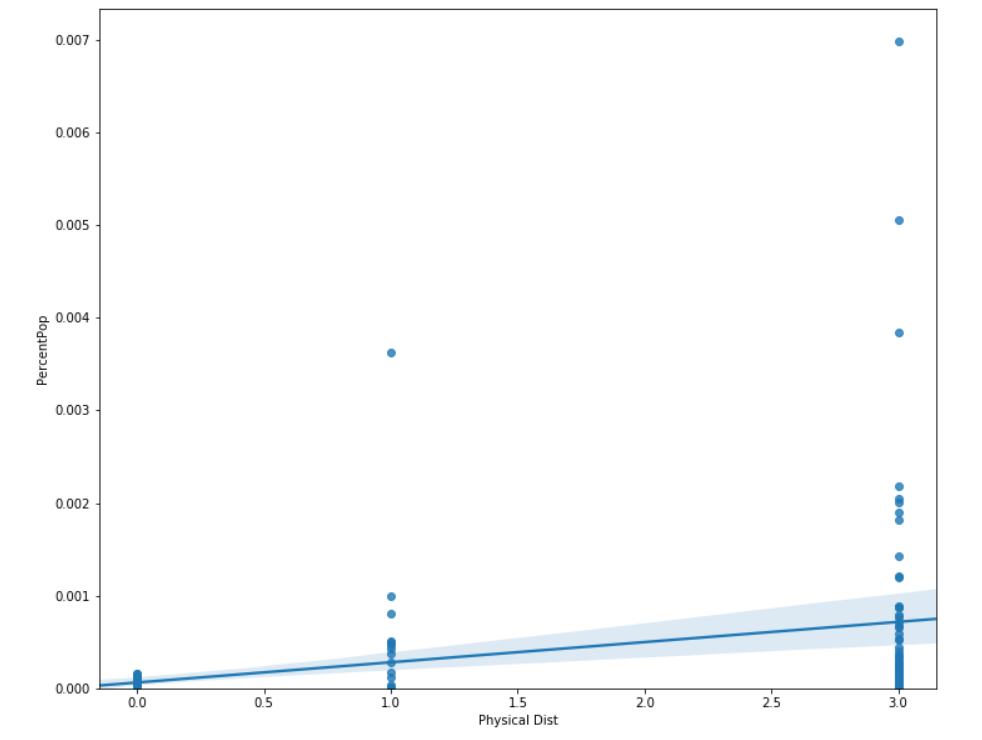


Figure 0=Unknown, 1= No Phys Distancing, 3=Phys Dist Implemented

There is no correlation between population density and the rate of infection.

